Supplemental Document

Biomedical Optics EXPRESS

Deep learning network for parallel self-denoising and segmentation in visible light optical coherence tomography of the human retina: supplement

TIANYI YE,[1] JINGYU WANG,[2] AND JI YI[1,2,*]

[1] *Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, 21231, USA*
[2] *Department of Ophthalmology, Johns Hopkins University, Baltimore, MD, 21231, USA*
[*] *jiyi@jhu.edu*

Supplement DOI: https://doi.org/10.6084/m9.figshare.24408280

Parent Article DOI: https://doi.org/10.1364/BOE.501848

A DEEP LEARNING NETWORK FOR PARALLEL SELF-DENOISING AND SEGMENTATION IN VISIBLE LIGHT OPTICAL COHERENCE TOMOGRAPHY OF HUMAN RETINA: SUPPLEMENTAL MATERIALS

1. Loss curves and explanation

Our data was meticulously split into training, validation, and test sets, maintaining an appropriate ratio. We ensured that they had a similar data distribution, encompassing most regions of the retina. During our experiments, we closely examined the loss curves and accuracies for both tasks, including Dice for segmentation and PSNR/SSIM for denoising.

Our experiments revealed that both training and validation loss curves demonstrated a generally steady decrease and eventual convergence (Fig. S1). The trend was similar across our experiments. We observed slight fluctuations in the validation loss during the initial stages of training for our segmentation model (Fig. **S1a**). These fluctuations can likely be attributed to 1) The complexity of the model we've employed may lead to initial overfitting of the training data, causing temporary increases in the validation loss. However, as training proceeds, the model tends to generalize better, resulting in improved performance; 2) The validation set we're using may inherently contain some variability, which can manifest as minor fluctuations in the loss. This effect is more pronounced when the validation set is relatively small.

In addition, when separately considering denoising, we've noticed that the difference between the validation loss and the training loss is not substantially significant (Fig. **S1b**). This does not indicate underfitting, as underfitting typically results in a considerably large gap between the validation loss and the training loss. In this case, we speculate that the observed behavior may be attributed to 1) The N2V loss employed during unsupervised training may have characteristics that contribute to this phenomenon. 2) The relatively small size of the validation set may not pose a significant challenge to the model compared to the augmented training set.

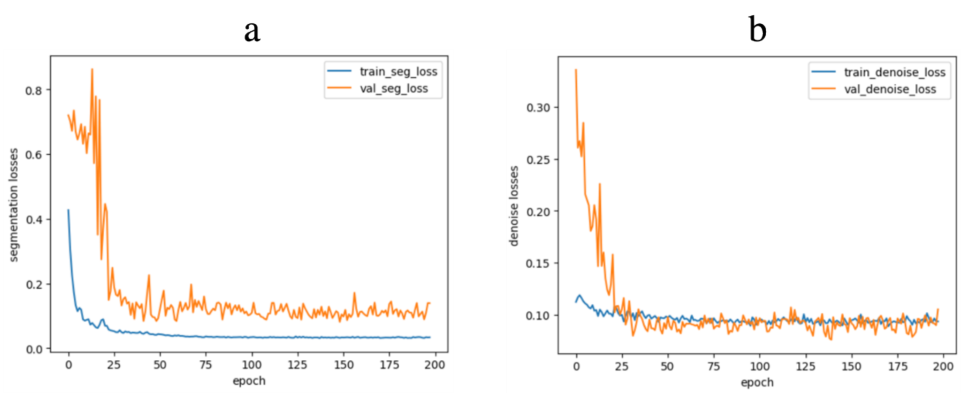

Fig S1. Exemplified training and validation curves for segmentation (a) and denoising (b), respectively.

2. Impact of vessel shadow artifact from segmentation.

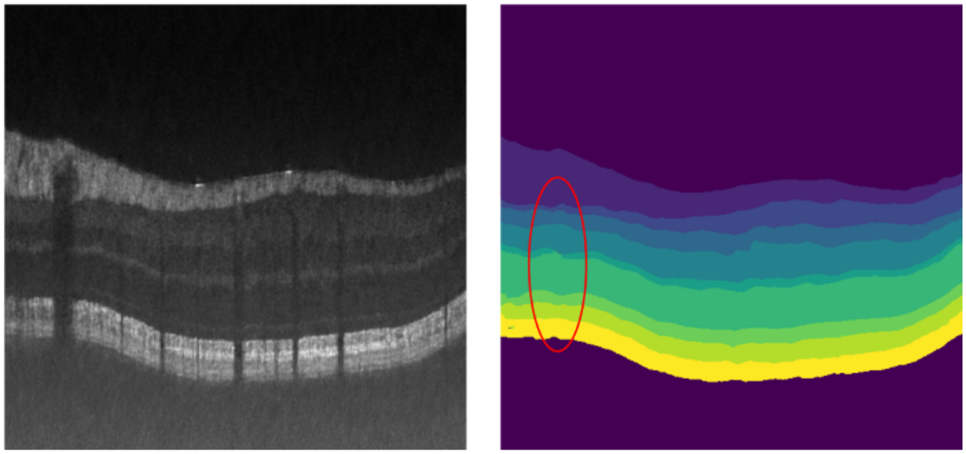

Fig S2. Impact of blood shadow artifact. The red ellipse indicates the area of a large blood vessel casting a wide shadow artifact.